Center for Independent Experts (CIE)
External Independent Peer Review Report

Review of the North Atlantic Right Whale Decision Support Tool

Dr. Julie van der Hoop

January 2020

Executive Summary

Under the MMPA/ESA, serious injury and mortality related to human activities is to remain below a specified level of Potential Biological Removal (PBR), established on a per-stock basis. For the North Atlantic right whale (NARW), PBR is currently 0.9 serious injuries and mortalities per year (or no more than 4 mortalities in 5 years). To reduce mortalities and serious injuries to below PBR, the Atlantic Large Whale Take-Reduction Team (ALWTRT, or TRT) would like to evaluate different methods by which to reduce the risk of lethal entanglement of NARW in fishing gear that include both closures and gear modifications, on the same currency. This requires 1) the definition of a target risk reduction (i.e., what level of risk reduction is needed to reduce serious injuries and mortalities to below PBR?) and 2) a method to evaluate the magnitude of risk reduction achieved by different proposals with diverse management methods (i.e., what change or combination of changes will reduce the risk of lethal entanglement to meet the established target?).

NMFS has therefore developed a Decision Support Tool (DST) which provides a framework for transparency, to identify realistic management choices, and to allow for ownership and agency over proposed, evaluated, potential, and implemented management strategies. The DST combines a model of the density of fishing gear, the density of whales, and the threat the gear poses to the whales, to develop a risk landscape, wherein management options are compared in their effectiveness in reducing lethal entanglement risk.

The gear and whale input models are effective in combining different data sources and for attempting to characterize complex point-estimates in geographic space. However, they have serious limitations in that the gear model is highly derived, and is largely opaque, with no estimates of uncertainty or variability. The whale model is informed solely by systematic survey effort, and as such is biased low in many areas, but does not leverage existing datasets to inform the effect of known biases. The whale model results must 1) be checked for agreement with these existing alternative data sources, 2) the areas of disagreement must be highlighted (and the reasons why determined), 3) the density model adjusted so that the density in these areas reflect the known presence of whales based on other survey and detection methods, and 4) the sensitivity of the DST output in these cases must be evaluated.

The threat model attempts to estimate the lethality of an entanglement should it occur, though data are very limited to inform this process. As such, the threat model is in a strong development phase, but is not ready for full application in decision-support as present.

The DST as a whole is strong in its framework and approach, but is limited by its lack of sensitivity analysis. The accuracy and sensitivity of the predictions are critical to resolve as they form the basis for mitigation strategies, environmental management and conservation.

Decision-making need not be reliant on the DST in its current development stage. We already know that any line in the water column poses lethal entanglement risk to whales, and that there is latent effort in the fishery. A decision could be made while the DST continues in its development, ready to inform the next phase of rulemaking in 3 years.

The unidentified serious injury and mortality 50/50 apportionment across the US and Canada, which determines the degree to which risk needs to be reduced, does not require any additional data or delay.

I understand that prompt action is required for the North Atlantic right whale. I also understand that the grounds upon which these decisions are to be made needs to be strong. I believe there is a way to achieve both.

Background

Under the MMPA/ESA, serious injury and mortality related to human activities is to remain below a specified level of Potential Biological Removal (PBR), established on a per-stock basis. For the North Atlantic right whale (NARW), PBR is currently 0.9 serious injuries and mortalities per year (or no more than 4 mortalities in 5 years). Human-caused serious injuries and mortalities of NARW have exceeded PBR annually since 2000 for all but one year, and incidental bycatch/entanglement in fishing gear has been identified as the primary cause of death when determined (Sharp et al., 2019).

When serious injury and mortality occur at a rate greater than PBR due to incidental takes in commercial fisheries, a Take-Reduction Team is established to provide stakeholder involvement in the development of consensus-based recommendations for take-reduction measures. The Atlantic Large Whale Take Reduction Team was established in 1996 with the purpose of developing a plan for reducing the incidental take of right, humpback, and fin whales, reducing serious injuries and mortalities to below PBR.

To reduce mortalities and serious injuries to below PBR, the TRT would like to evaluate different methods by which to reduce the risk of lethal entanglement of NARW in fishing gear that include both closures and gear modifications, on the same currency. This requires 1) the definition of a target risk reduction (i.e., what level of risk reduction is needed to reduce serious injuries and mortalities to below PBR?) and 2) a method to evaluate the magnitude of risk reduction achieved by different proposals with diverse management methods (i.e., what change or combination of changes will reduce the risk of lethal entanglement to meet the established target?). The approach further allowed for states and stakeholder groups to craft their own measures that would reduce risk to reach the target in a given area; these proposed measures could then be combined, where together the components would achieve a 60-80% risk reduction for the entire domain. NMFS has therefore developed a Decision Support Tool (DST) that would support this type of assessment for a mix of measures.

Decision-support tools are popular approaches to determine an optimal or best approach, especially in complex or dynamic situations with diverse stakeholders and multiple solutions (Bardos et al., 2001). By definition, decision-support tools have six characteristics (Geoffrion, 1983):

i. explicit design to solve ill-structured problems
ii. easy-to-use and powerful user interface
iii. ability to combine analytical models with data
iv. ability to explore the solution space by building alternatives

v. capability of supporting a variety of decision-making styles
vi. allowing interactive and recursive problem-solving.

In this way, decision-support tools provide a framework for transparency, identify realistic management choices, and allow for ownership and agency over proposed, evaluated, potential, and implemented management strategies. Further, it allows agencies and stakeholders to weigh the costs of collecting and/or analyzing new data against the value of additional data in arriving at a more robust decision (Geoffrion 1983). They allow for uncertainty to be resolved through exploring the effects of alternative conceptual models and parameter choices on the decision, and serve to make the decision-making process transparent, documented, reproducible, robust and provide a coherent framework to explore the options available (Geoffrion 1983).

The Decision Support Tool (DST) developed by NMFS combines a model of the density of fishing gear, the density of whales, and the threat the gear poses to the whales, to develop a risk landscape. The DST focuses on the fishery with the greatest contribution to the number of endlines in the water – the northeast region trap and pot fisheries. The tool evaluates how the following management measures (alone, in part, or in combination) would change risk to whales across the region:

1. Seasonal Closures, either with gear removed from the water or allowed to redistribute
2. Trap Reductions
3. Regulations in Trawl Length
4. Regulations in vertical line characteristics
5. Implementation of ropeless or timed-release technology.

Description of the Individual Reviewer's Role in the Review Activities

As a reviewer, I have read and reviewed materials provided prior to the meeting, asked for additional clarifying and up-to-date information from NOAA and other right whale colleagues, traveled to the meeting, and volunteered to facilitate the meeting, for which no chair had been pre-arranged. I installed the DST and ran scenarios, asked follow-up questions, and received additional clarifying information after the conclusion of the meeting and while preparing this report. I had previously taken part in an hour-long call regarding sub-lethal effects of entanglement to inform the DST May 30, 2019; however, I was not aware of DST structure, data inputs, assumptions, or outputs.

Summary of Findings

TOR 1 Evaluate the data inputs used in the decision support tool.

The inputs to the Decision Support Tool are 1) a vertical line model prepared by IEc, representing the density of vertical lines from fixed-gear fisheries; 2) a whale density estimate model prepared by Duke University; 3) a severity model that estimates the risk of serious injury or mortality as a function of rope breaking strength.

The DST then serves to combine these three models to estimate risk, defined by the DST as the product of gear threat per endline, density of endlines, and density of whales. The DST is designed to allow for user inputs to create scenario situations to assess how risk is reduced by introducing management measures including seasonal closures, trap reductions, trawl-length regulations, gear characteristics, and buoy-less fishing.

1) **The Vertical Line Model** is intended to provide the best estimates of the seasonal distribution of commercial fishing gear off the US Atlantic coast. It provides the "gear layer" to the DST. The Vertical Line Model outputs the density of lobster traps throughout the domain of the model at a 1 NM spatial and monthly resolution. These data, along with co-located data on trawl length, are inputs to the DST.

<u>Strengths:</u>

- The Vertical Line Model is designed with sufficient flexibility so as to include and leverage very different and disparate data sources that characterize the US Atlantic fishing industries.
- The developers of the Vertical Line Model have attempted to validate some of their estimates, though their documentation does not provide the details of this validation, and it is only performed for the Massachusetts' 14 inshore statistical reporting areas (SRAs). Though the annual vertical line use is in agreement with reported vertical line use, the agreement varied by month and SRA (IEc 2014 VL Model Documentation). Note that the details on this validation procedure, and quantitative measures of the agreement and divergence, and effect of season have not been provided.
- An intention of the Vertical Line Model's developers is to involve the TRT in model design, data collection, and informing various assumptions. It appears that many TRT members and representatives of state fishery management agencies are involved in the development and improvement of various aspects and assumptions of the Vertical Line Model. This teamwork is appreciated in producing a tool that is transparent to all stakeholders.

Concerns:

- The IEc Vertical Line Model has been under development since 2005. This continued development has applied different methods to different time periods; as such, there is no comparable time series over which to test specific assumptions regarding the behaviour of the fishery. For example, it would be useful to test how variable the fishery is from year to year. This is not possible given structural changes between model iterations and data updates. If the Vertical Line Model is to be chosen as the best representation of the truth, IEc should be able to apply one consistent method to all historical data, so as to quantify the variability of the fishery and how that variability will impact the accuracy of future predictions. This would also serve to test for the effect of certain structural changes or assumptions as the model has developed, as data vs. handling assumptions currently confound such an analysis.
- Resolution: The Vertical Line Model uses data at 1 and 10 NM resolution, but also applies other values uniformly across statistical areas. The resolution of the Vertical Line Model is therefore the statistical area. Documentation for the review, NMFS (2019b), even states that "the geographic precision of the model's presentation … may be overstated" and that, as data are assigned to individual grid cells, the output implies "a higher degree of geographic precision… than the underlying data warrant." Still, the output of the Vertical Line Model is the number of traps per 1 NM cell per month, reported to 10^0. What are the actual data requirements to interpret the model at a certain level? (e.g., annual vs seasonal, 10-minute vs 1-minute)?
- Complete lack of sensitivity/uncertainty: The IEc model requires extensive data transformations and values are highly derived. Many of these are valid given the disparate nature of datasets; however, they are poorly described (the Panel was provided a description for the 2014 methods). Further, IEc reports no quantified uncertainties on the final outputs and overestimates the geographic precision in characterizing gear densities based on the scale of the underlying data. The gear model entirely lacks a sensitivity analysis to quantify the effect of various assumptions on the final output.
- Lack of response to 2012 CIE reviewer comments: The Vertical Line Model was subject to a CIE Peer Review in 2012. All three reviewers noted the absence of explicit consideration of uncertainty in the model. I see no improvement on this front in the last 7 years. For example, the Vertical Line Model uses the average location from VTR; what is the sensitivity of the output to e.g., geographic misallocation of 10% of the reports? Does this completely alter our interpretation of spatial or seasonal gear density, or is the effect muffled by other factors in the derivation?

- The CIE Reviewers were asked to provide a critique of the NMFS review process, including suggestions for improvements of both process and products. The Vertical Line Model was subject to the NMFS/CIE review process in 2012; however, the major and consistent points raised by the reviewers have not been addressed in any way. This makes me question the use and value of a review process, if the subjects of the review do not implement proposed changes, or at least provide clear response to the requests of reviewers detailing why they are unable to implement proposed changes. This is one of the most key steps in a peer review process: refining the product, responding to reviewers, clearly stating how and where the product has changed in response certain feedback, and respectfully declining to incorporate other points of feedback.
- <u>Methods and assumptions are poorly defined and described:</u> Despite the intention of providing transparency, the IEc model description is opaque. The documentation for the most recent iteration of the Vertical Line Model provided to NEFSC is not complete. The reviewers were provided documentation from 2014, which provides some information on general methods, but does not inform the Panel of detailed methods, changes from the 2014 to 2019 versions, or the outputs, assumptions (and effects of those assumptions), and methods that truly underpin the DST.

<u>Recommendations:</u>

I recommend that the 2012 CIE Reviewer Reports of the IEc Vertical Line Model be revisited, and that IEc draft responses to the recommendations. Many points raised by those reviewers are still valid today. This current version of the model lacks documentation, and the derivations of the output metrics remain opaque, and the model itself demonstrates no attempt to identify high or low estimates, or any sort of sensitivity analysis.

2) **The whale density model** is a density surface model initially developed for the purposes of the US Navy (Roberts et al., 2016). The DST is currently using v8 of the whale model, which includes right whale sightings from strategic aerial and vessel surveys from 1998-2016. The density model accounts for detection probability, including with distance, sea state, regional differences in whale behaviour and group size, as well as differences in detection probabilities between survey platforms. Density surface modeling (DSM) then continues to correlate available oceanographic and bathymetric data including climatological estimates, and uses region and season-specific spatial models to estimate the density of NARW in areas

with low or no survey effort. The output of the whale DSM is the number of whales per 10 km grid cell per month.

Strengths

- The whale DSM approach is rigorous in that it incorporates available systematic surveys from multiple platforms, applies methods from an established research approach (distance sampling) while also incorporating select oceanographic variables in an attempt to improve estimates for areas where survey data are limited.
- The DSM approach provides estimates of uncertainty and the Coefficient of Variation. This is critical in informing 1) where additional effort should be directed; 2) where interpretation should be limited based on the high level of uncertainty; and 3) where there is high confidence in the spatial and temporal density estimates. This uncertainty or CV map contextualizes the output of the whale model and is ready for input to DST; however, this uncertainty is not currently propagated forward as an input to the DST.

Concerns:

- Exclusion of available systematic survey datasets: Systematic surveys by NEAq, CCS, NLPSC and DFO are not included in v7 or v8 of the model. These would be extremely important to include given the magnitude of effort for these three contributors. Adding these and other more recent data (v9; 1998-2018) provides a worthwhile point to assess the sensitivity of the whale model to the addition of a new "year" of sightings and effort. Does the addition or removal of a year affect the output by 1%? 10%? 50%?
- Exclusion of the majority of right whale sightings effort: Data inputs for the DSM approach must be compatible with systematic survey design. Survey effort must be directed at predefined and random times, not towards recent knowledge of animal presence or absence. There are many surveys that fit this description that sight NARW. However, the majority of survey effort for NARW is directed to inform specific studies e.g. mark-recapture (Pace et al., 2017; Crowe et al., 2019), photo-ID (Schick et al., 2013; Rolland et al., 2016; Pettis et al., 2017), biopsy (Malik et al., 1999; Frasier et al., 2007), tagging (Nowacek et al., 2004; Baumgartner and Mate, 2005; Parks et al., 2011; van der Hoop et al., 2019), etc. These directed surveys and effort-corrected sightings are not included in the DSM, yet they would serve to improve or evaluate times and areas where the DSM predicts zero whales, but where effort-corrected sightings data suggest otherwise.

- <u>Lack of coverage in inshore waters:</u> Related to the above, the DST documentation states that the whale model lacks coverage in inshore waters (Fig 4.8a in DST documentation). The documentation additionally states that NARW are "exceedingly rare in the missing areas" – but based on what data? Targeted surveys, opportunistic sightings data (Knowlton et al. 2019), satellite telemetry (Baumgartner and Mate, 2005) and acoustic data suggest that right whales are detected in inshore areas in ME and other states in the summer and fall months. The statement that NARW are "exceedingly rare" in these areas with "minimal effect on scenario testing" is entirely due to the data sources used and those that are ignored. Further, this "effect on scenario testing" is entirely unquantified.
- <u>Developers' request for "more data":</u> It is concerning that the developers request more data to inform their models yet they are 1) lagging to incorporate data by 1-2 years, and 2) are ignoring complementary data sources and datasets that inform distribution and residency that are not being used.
- <u>Model Accuracy:</u> Accuracy of the density model based on validation with e.g., existing, alternative datasets mentioned above, has not been completed. This accuracy assessment would highlight areas where density estimates and Sightings per unit effort estimates from non-standard surveys or acoustic data suggest very different levels of whale occupancy, especially at certain times of year. These highlighted areas and times would be regions to consider for improvement of the model, especially if gear density is high in those regions, i.e., where under-predicted whale density in that area would have a considerable effect on our estimates of encounter probability in that same area and time of year. The reviewer understands the challenges of integrating different datasets, and does not recommend that a full framework to integrate acoustics into density estimation be pursued. However, 1) the density model results must be checked for agreement with these existing alternative data sources, 2) the areas of disagreement must be highlighted (and the reasons why determined), 3) the density model adjusted so that the density in these areas reflect the known presence of whales based on other survey and detection methods, and 4) the sensitivity of the DST output in these cases must be evaluated.
- <u>Acknowledgement of limitations:</u> At any given time of the year, we don't know where at least 75% of the population is. Additionally, fewer than half of individuals in the population can be accounted for in any given year. The whale model (as published; Roberts et al. 2016) acknowledges the seasonality in CV (0.45 from Nov-Feb) but these values are not propagated through to the DST. Other fields in spatial ecology and oceanography provide excellent inspiration

for how to include and propagate uncertainties from point estimates through to uncertainties in distributions (e.g., Meyer et al., 2016; St. John Glew et al., 2019).

- Aggregate monthly history: The whale density models are created on a monthly basis from aggregate survey effort from 1998 through 2016. In this way, they do not represent the density of whales in a given area in a particular year, nor will they provide accurate or precise predictions. Additionally, the DSM treats sightings as independent and identically distributed— which they are not. In fact, they represent multiple sightings of the same ~600 individuals over 18 years. It does not discuss how this important assumption limits inference from the density model from a risk-estimation standpoint.
- Model fit: Because the density model is trained on a subset of available survey data, this set of biased coordinates may link whale occurrence to false environmental conditions. The selected oceanographic features may describe the systematic survey observations, but not observations captured in the data sources mentioned above (acoustic, other surveys, opportunistic sightings, and telemetry data). This is clearly seen in e.g. the mid-Atlantic in the winter months, where few sightings and limited survey effort mean that only a single covariate (distance to shore) is used to inform the spatial estimates, and where the model artefact is evident on the maps. The Model is then extrapolating to areas where there are no data (e.g., inshore Maine) with oceanographic variables that are likely poor predictors of NARW distribution.
- Selected habitat features: The single most important habitat feature for the NARW in foraging months is the occurrence of concentrated patches of copepod prey (Murison and Gaskin, 1989; Mayo and Goldman, 1992; Wishner et al., 1995; Mayo et al., 2001; Baumgartner and Mate, 2003) which is largely driven by water-mass structure (Davies et al., 2014; Davies et al., 2015). Water mass structure, however, is not included in the whale density model. The DSM does include biological factors (chlorophyll a derived from satellite imagery, net primary production from satellite imagery and the VGPM ocean model, zooplankton production and biomass from the SEAPODYM ocean model, and epipelagic micronekton production and biomass from the SEAPODYM model); however, the resolution of these layers is at either 9 km or 0.25° – perhaps too coarse to be relevant to fit to NARW sightings from systematic survey effort. Given the importance in water mass structure (recently used to identify potential habitat in the Gulf of St Lawrence that found a previously unknown aggregation of NARW), water mass structure may serve a fruitful avenue as a correlate for at least foraging areas (Davies and Brillant, 2019).
- Bias in g(0) estimation: The distance-sampling method requires an estimate of g(0), the probability of detecting an animal on the transect line. Roberts et al.

(2016) turned to the published literature to inform estimates of g(0) based on geographic region. This is appropriate, given that right whales show different foraging patterns in e.g. Cape Cod Bay versus deep basin habitats (Mayo and Marx, 1990; Baumgartner et al., 2017). However, these published manuscripts present a bias in behaviour, depending on their focus. In basin habitats, g(0) is estimated from dive data reported by Baumgartner and Mate 2003, where the surface interval is 3.13 min and dive is 12.17 min. This is inappropriate, as the 3.13 min is the surface interval following foraging dives (12.17 min average duration), not the amount of time that right whales are at the surface on average. Right whales do not continuously and consistently forage, even in their most productive habitats. In the same paper by Baumgartner and Mate (2003), figure 3 illustrates how right whales spend their time traveling, socializing, and searching for food, in addition to foraging. Even in their most productive habitats, right whales will go without foraging for extended periods of time (>2 hours) (van der Hoop et al., 2019). This behavioural time budget is critical to acknowledge in the calculation of g(0) as the current method overestimates the time spent at depth. Sensitivity of the DSM to g(0) has not been tested.

Recommendations:

- Overall, high-density areas predicted by the DSM concur with other data sources. However, there is disagreement where systematic survey effort is low but directed survey, opportunistic, telemetry and acoustic data are available.
- I recommend that the authors and the DST developers acknowledge the limitations of the DSM model approach and supplement it with other available data. Are zero cells actually zero? This is a critical assumption to test, and one to which sensitivity of the DST output is also untested.
- The uncertainties derived by the DSM approach should be input to the DST and propagated through. Running the DST on the bounds of the density estimates (low and high) would also suggest the degree to which results are consistent, or areas where confidence regarding risk assessment with this combination of models is high.

3) **The gear threat model** attempts to quantify how gear configuration contributes to the probability of serious injury or mortality should entanglement occur.

This model is largely based on the severity of injury based on line diameter or line strength. A study (Knowlton et al., 2016) has shown that high breaking-strength rope more often results in more severe injuries (note: a definition separate from Serious Injury). The gear threat model attempts to assign a likelihood of mortality given rope strength, which provides another mitigation strategy, and reflects the important tradeoff of trawling up (fewer lines, but more traps per line, which requires stronger rope) or reducing breaking strength (lighter line, easier for a whale to break, but can support fewer traps, so more total lines in the water column).

Strengths:

- Purpose: This model layer is needed to achieve the goal of providing some quantitative analysis of how changes in gear configurations (rope strength, trawl length, buoyless fishing, etc.) each contribute to decreasing risk to whales. Therefore, the DST achieves the initial goal of creating a mechanism that allows for direct comparison of different management solutions to reduce lethal entanglement risk.
- A reexamination of entanglement outcomes: I appreciate the breakdown of gear weight vs rope strength with respect to entanglement outcome. I believe this thought model is useful for ideation and design goals, given that it identifies the outcomes of the decisions that can be implemented. This is likely useful at the level of governance rather than the DST, whereby NMFS has to decide: would they rather have fewer whales tethered, where the injury is near-immediately lethal but rarely observed; or, have individuals break free, trail the gear, and die 6-12 months later due to energetic and health consequences. Again, this is not for the DST to consider, but rather for the governing body to decide how they would rather proceed in managing, mitigating, and allowing takes.

Concerns:

- Focus on a single factor affecting entanglement outcome: The DST threat model focuses entirely on rope strength. While gear removed from whales and subsequently strength tested suggests that injury severity increases with rope strength (Knowlton et al., 2016), it is only one factor of the many that contribute to serious injury or mortality due to entanglement in fishing gear. Serious Injury criteria have been clearly defined for entanglement (NMFS, 2012) and these definitions have been further developed for entanglement in large whales (Moore et al., 2013). The focus is commended (as it can be useful at times) but given the lack of fit of the model to the data, other factors should be considered in a multivariate approach. It is important to note that based on the data, the probability of lethality is effectively 1 at all rope breaking strengths.
- Entanglement Process: Related to the above, we know little about how a co-occurrence becomes an entanglement. While we do know the characteristics of the gear set in certain regions, and the characteristics of different sets of gear removed from whales, this remains too n-limited (<4% all entanglements) to rely on to solely inform a severity model. Even if we were to remove all gear from

whales, we would not know what proportion of the original gear it represents (i.e., what was the original entangling set versus what remained on the whale) and how the characteristics of the gear set affect the entanglement likelihood and/or the outcome.

The challenge is that we know little about how a co-occurrence becomes an entanglement. We know the characteristics of the gear set in certain regions, and some characteristics of different sets of gear removed from whales; however, this sample size is simply too small at ~4% all entanglements, to rely on to solely inform a severity model. From 1998-2017, there have been 1544 entanglement interactions; only 122 (7.9%) with attached gear. Only for a subset of those 7.9% is gear retrieved or removed, and only half the gear removed from whales can be identified to a fishery (Johnson et al. 2005). From 2010-2017, gear was retrieved from only 4% of confirmed entanglement cases (Knowlton et al. 2019 NARWC). "Given that actually observed entanglements are far less frequent than entanglement scarring would indicate, it follows that opportunities to sufficiently examine and to determine the origin of the gear is even less frequent; i.e., associating the prevalence of gear type with entanglement will remain n-limited for some time to come." (Brilliant et al. 2017)

- <u>Overwhelming sensitivity to threat:</u> Whatever threat model applied will largely affect the scenario-testing results and therefore the management outcome. The certainty and confidence in the threat model is therefore paramount in the establishment of the DST as a whole.

4) **DST treatment of inputs:** The DST itself imports the gear information (initial density of traps by location and month); implements rules for scenario testing (general trap removals, implementation of new trap caps, spatial closures, vertical line characteristics, and implementation of ropeless or timed-release technology); estimates the density of trawls based on regulations, reporting, and returns, dependent on state; and estimates the number of endlines per trawl by regulations by state, to estimate the density of endlines across the region, by month. The exception is LMA3 where observer data (primarily from 2014-2015), landings, federal VTRs, and bathymetry are leveraged to provide more spatially explicit estimates of gear density. Combined with the whale density estimates at a 10 km resolution, the DST estimates the co-occurrence and the potential risk to right whales posed by vertical lines used particularly in the lobster fishery, as it accounts for most vertical lines in use. The model incorporates relevant new information as it becomes available.

Strengths

- LMA3 line density estimates: The approach for estimating seasonal trap densities in LMA3 is highly derived but is an interesting direction. This method at least informs the distribution of fishing effort around e.g. lobster habitat, instead of assuming uniform effort across a management area. I wonder if this approach could be used to at least test the effect of the sensitivity of the DST as a whole to IEc's assumption of uniform effort within a statistical area.
- Ability to accommodate all management types: The DST is flexible in that its submodules can accommodate closures, reduction in effort, caps, and buoyless fishing. The language around closure to buoylines, not to lobster harvest, is critical for supporting the development of ropeless fishing technologies and implementing them alongside the other management strategies.

Concerns:

- Buoyless fishing line-reduction estimates: The DST documentation notes that estimates for the reduction in the amount of line achieved by buoyless fishing was informed by a TRT opinion poll: an 88% reduction in lines for acoustic releases and a 52% and 46% reduction in lines for timed releases inside and outside of 12 miles offshore respectively. The DST documentation additionally states that exact numbers for buoyless fishing are hard to obtain "as this is still largely untested technology." I am concerned with two aspects of these statements and the approach: a) describing ropeless technology as being "largely untested", and b) the availability of at-sea testing data that can inform the actual reduction in vertical line associated with acoustic release technology.
  a) The Ropeless Consortium has held annual meetings since 2018, where stakeholders and interested parties have presented on at-sea testing, technology development, market size, and policy avenues. At the 2019 meeting, at least seven manufacturers provided updates on their product, many including at-sea trials in a diverse set of trap/pot fisheries in different conditions (e.g., Gulf of St Lawrence Snow Crab, Massachusetts Lobster, offshore lobster). NOAA and NEFSC are additionally conducting collaborative buoyless research with many manufacturers (https://ropeless.org/2019-annual-meeting/).
  b) The above at-sea tests would be a more suitable source for estimating the vertical-line reduction achieved by buoyless gear. For example, NEFSC testing shows that buoys are observed within 1 minute of triggering the acoustic release, and are then manned throughout the repacking / hauling / redeployment process. The released line is only in the water column for the 15-20 minutes that the gear is being actively hauled. If gear is hauled once every 5 days, the line is

in the water for only two hours per month, compared to traditional gear, soaking unattended for 720 hours. This back-of-the-envelope reduction in soak time is much greater than 88% for acoustic releases (It is a 99.7% reduction). Studies and gear tests, where data do exist, would better inform the effect of buoyless fishing as a mitigation technique, instead of mean values from a poll.

- Focus on a single portion of the fishery: Limiting the scope to lobster and crab fishing only simplifies the analysis, but eliminates the opportunity to compare and contrast with other fisheries. On the technical side, I understand the decision to focus on the largest contributor in order to have a large impact with a single decision. On the stakeholder side, this may result in that particular group feeling targeted unfairly.
- Sensitivity analysis: The DST entirely lacks an analysis that describes the resulting change in the output based on small changes to the input values or functions within the DST framework. Understanding the sensitivity of the conceptual model to parameter choices is necessary before moving forward in using the model as a scenario test. The tool must be sufficiently robust to support the types of decisions being made.
- Propagation of uncertainties: Various submodules of the DST include statistical modeling that results in models to obtain the mean and variance of specific parameter estimates. The mean is propagated through to the next level or submodule, but the variance is left behind. Some submodules include robust approaches (e.g., bootstrapping) but this approach remains only within each submodule. What is the uncertainty for any of these parameters when resolved with the DST? Even just based on a sequence of assumptions, are the resulting threat and risk estimates too high or low?
- Redistribution of traps after closure: An option in the DST is to implement a closure, but redistribute the effort elsewhere. There was concern from the industry that the behaviour prescribed in the model (linear cost function with redistribution value based on high effort indicative of productive lobster habitat) does not reflect true behaviour of the fishery. While this is a good first approach, the redistribution model does not realistically portray the cost relative to the value of deploying gear where catch is high. Note the recent effort on the effect of closures in the Canadian fishery (Cole and Brillant, 2019).

TOR1 Overall Requirements:

- A fully laid-out series of assumptions and rationales: this is required to explain why certain choices were made and their impact on the final outcome (e.g., conservative estimates, introduced bias, sensitivity to x variable). As a reviewer, I am not as much concerned with the data themselves as I understand the limitations and the need to act; however, I am concerned about how the data are handled, what assumptions were made, and how the handling and assumptions are poorly described and untested for their outcome. It is alright to have data gaps and assumptions to fill them, as long as they are well articulated and framed as the "best-available" techniques and data available.

- Quantified overlap of inputs: Measuring and evaluating the overlap of whale strategic survey effort and of fishing effort data is critical to assess the gaps and where the model is better or more poorly informed.
- Sensitivity analysis of the DST inputs and submodules and propagation of uncertainties are required prior to using this tool to make decisions.
- An established scale: The whale density model has a native resolution of 10km pixels, others are at 1 min grid, 10 min grid, and stat area.
- Limit the requirement to predict whale presence: Given our inability to locate the majority of the right whale population at any given time of year, and to predict when they might arrive in a certain location, it is likely cautious to reduce the dependency of the risk-reduction framework to predicting whale presence on a small spatial or temporal scale. Testing broad scenarios is likely warranted given the coarse resolution and aggregate estimates of gear and whale densities.

TOR 2 Evaluate the data outputs produced by the decision support tool.

A complete list of the outputs from the DST are provided in Appendix 4.

By definition, outputs of a DST should be “scientifically defensible, although not necessarily accurate, predictions of the system behaviour” (Li, 2019). Currently, the outputs of the DST (and its inputs or underlying models) are not truly scientifically defensible (i.e., they lack sensitivity and uncertainty analyses), nor are they likely to accurately predict system behaviour based on the selective use of and lag in whale data, and the assumptions and inconsistent time-series of the gear data.

Strengths

- Configuration Details: The outputs are thoughtful in that each report contains specific configuration and input settings. This level of documentation allows for reproducibility of the baseline and scenario runs.

Concerns:

- Precision and accuracy of the estimates: The values reported in output tables give the reader a sense of precision or accuracy that is not the case and could be misleading to the public or the end-users of the tool. Nuances in the model are quickly lost once a final risk-reduction score is produced. The reporting should be to the level of precision that is justifiable from the inputs (van der Bles et al., 2019). Because of how the DST is likely to be used, the degree of certainty must be provided alongside all estimates: automatically rounding up or providing some metric of variability around e.g. mean values.
- Uncertainty: Having a measure of uncertainty is important – in fact, it can help determine the major contributing sources, those that have the greatest effect on the output, and therefore where to focus additional efforts (or whether additional efforts would or would not truly change the end result). One of the hallmarks of DSTs is that they can help weigh the costs of collecting and/or analyzing new data against the value of additional data in arriving at a more robust decision (Geoffrion 1983). Is variance too high to deal with, or is it suitable for decision-making?
- Definition and units of risk: The terms and units used in the DST submodels and its output may be slightly different to standard definitions. ISO 31000 defines risks as sets of triples: (1) a scenario (i.e., a hypothetical future event or set of events), (2) the likelihood of the scenario occurring, and (3) the consequences of the scenario (also referred to as the Kaplan-Garrick definition). Using such a standard definition, also based in probability with units of 0 to 1, puts the results in context with the public's general perception of probability and risk (or at least doesn't go against the major efforts in attempting to standardize to increase public perception and understanding). Standardizing risk units 0 to 1 also puts results in context with the previous literature in the area of risk posed by fishing and shipping to whales (Vanderlaan et al., 2008; Vanderlaan et al., 2011; Carr et al., 2018).
- Risk-reduction "credit" for immediate action: Because the whale model lags behind current observations, closures within a year of an observed large aggregation of whales in an area with gear (e.g., south of Nantucket) could not be included in the risk-reduction framework presented here. Because the whale and fishery data are in aggregate across years, and that neither whale nor vertical line model have demonstrated the stability of their products year-to-year (i.e., their sensitivity to annual fluctuations in whale observations and fishing effort), the ability for changes made based on changes in whale sightings and fishing will not be accurately reflected by the current framework.

Recommendations

- Residual plots for scenario testing. One of the current outputs of the DST is a set of maps for both configuration and scenario runs. These map sets have 12 panels, one for each month, to display the spatial distribution of e.g. trap density, trawl length, line density, line strength, threat, and risk over the course of the year. It remains difficult to see the effect of the scenario, compared to the default baseline condition, even comparing these maps side-by-side. A residual plot on a log scale, with a diverging colorscale (e.g., RdBu in RColorBrewer) would more clearly show areas where the scenario causes increases and decreases of a certain output. Such a residual map would be an effective way of seeing how a scenario's mitigation decisions and geographic reach influence the spatial and temporal distribution of e.g. line density.
- We may simply not be ready to mitigate based on gear configuration, only on gear distribution and effort alone.

TOR 3 Comment on the appropriateness of using the decision support tool to evaluate relative entanglement risk to right whales; advise on the strengths and weaknesses of using it to compare management measures.

Strengths:

- Teamwork: the DST approach supports teamwork and involvement with all stakeholders. Industry members are able to run scenarios on their own, which supports ownership and creativity among all stakeholders.
- Transparency: The distribution of capacity and of "power" in designing, proposing, and testing scenarios is effective in allowing for stakeholders to truly engage in the process and understand the basis by which mitigation strategies will be designed and implemented.
- Framework: The Information - consequences - choice framework is especially useful for industry members to evaluate options and identify solutions *they* would like to implement on their own, instead of ideas being proposed to them that they may not deem feasible.

Concerns:

- Impact of assumptions: What would happen to the number of endlines if CPUE increases? Then how would that affect the outcome or output from the tool? It is inappropriate to interpret the output for management decisions without a quantified understanding of how assumptions and parameter values will affect the final results. The DST somewhat lacks the rigour to be sufficiently robust to support the types of decisions being considered. I believe that it can get there, but it is not there quite yet.

- The Threat Model: Following concerns from TOR 2, the threat model requires a lot of assumptions, is highly derived, and has poor fit to a single parameter that has been treated as the primary factor in entanglement outcome. This threat model has a massive impact on the final outcome, as it is multiplied by gear and whale density.
- Development in Tandem: It is understandable that all elements of the DST will continue to develop, especially in light of reviews from many directions. This simultaneous and continuous changes to the DST, its input models, and its submodels, does pose a real challenge in 1) keeping stakeholders updated and 2) maintaining trust. For example, especially given that the sensitivity of outputs has not been established, there can be concern that the measures determined to meet a given level of risk reduction in version N.n may be insufficient when used in version N.n+1. The change in risk reduction between versions N.n and N.n+1 cannot be anticipated (i.e., will it be 5% or 50% different?). The ongoing efforts to make iterative changes on submodels and inputs also creates a challenge for 3) evaluating the DST overall. Currently, the DST is integrating components that are under development and construction or reparameterization, in addition to adding new data. These structural changes in the model between time-frames (specifically IEc) complicate the consistency of the DST output.

Recommendations:

- Specific improvements in structure and rigour are needed to be sufficiently robust to support the types of decisions being considered.
- The DST documentation should be put in context with literature on risk and decision support tools so as to provide consistent definitions and to leverage a lot of work done on standards in this research area.

TOR 4 Provide research recommendations to improve the decision support tool.

- Sensitivity analysis: This is the central thesis of the report. As such, I don't believe it requires further elaboration.
- Incorporation of complementary datasets: See TOR 2.2 – whale model. This is essential to inform the areas where the whale DSM is known to be biased low – areas where strategic survey effort is low, but where targeted survey, acoustic, telemetry, and opportunistic data are available.

- Product improvement > product development: The 2012 CIE reviewers noted the same themes for the Vertical Line Model – test the effect of current assumptions, before going any further. This has clearly not been done. At this point, development needs to pause and improvement needs to be the focus. Build the architecture, make it strong, and then fill it with the updated information when it is robust and ready. It will then produce defensible results, with measured sensitivity, and estimates of uncertainty, with known limitations but also a defined level of confidence.

TOR 5 Evaluate whether the methods represent the best available scientific approach for apportioning human-caused mortality by country.

One of the key factors in determining the risk-reduction goal or target is the reduction of serious injuries and mortalities to meet PBR. However, one of the major challenges is agreement as to how to apportion serious injuries and mortalities in a transboundary stock. There is strong desire to assign "Canadian" and "US" mortalities, and for the US to reduce risk to a level required for US mortalities to meet PBR. Note that the abundance estimate that defines PBR is calculated for the entire NARW population, not just the US proportion of the population (GAMMS I, 1996).

The NARW is not the first transboundary "stock" where this question has been raised, and precedent is important. Trans-boundary species guidelines already exist. (Barlow et al. 1995) and state that: "In trans-boundary situations where a stock's range spans international boundaries or the boundary of the US EEZ, the best approach is to establish an international management agreement for the species. In the interim, if a stock is migratory and it is reasonable to do so, the fraction of time in US waters should be noted, and the PBR for US fisheries should be apportioned from the total PBR based on this fraction."

The challenge for NARW is that mortalities and serious injuries occur for a single stock in two countries, and that there is some latency in detection of injury and mortality, in which time animals can move between regions. Apportioning based on the location where the entangled whale is first seen and the gear identified is unacceptable for the following reasons:

- Few whales have gear on them: The NOAA briefing document to CIE Reviewers states that "since 2010, gear was only retrieved from 26% of the right whales that were killed or seriously injured. In 45% of the cases during that period, there was no gear on the whales, and in 29% of the cases the gear was not retrieved."

- Few entangled whales with gear are sighted within 60 days of a previous sighting: Only 6.5% of the entanglements documented from 2010-2015 from scarring events and seen carrying gear were observed within 60 days of their previous sighting (Knowlton et al. 2019 NARWC). Within these 60 days, subsequent sightings were split between being within the same region on the eastern seaboard, and having made long-distance movements to a completely different region.
- Ability to identify gear: Due to the known challenges in implementing an effective gear marking scheme (GAO, 2009), and the different volumes of gear found on and retrieved from entangled whales (relative to the original gear set), gear identification is a real challenge. From 1980-2017 there have been 1544 entanglement interactions; only 122 (7.9%) with attached gear. Only for a subset of those 7.9% is gear retrieved or removed, and only half the gear removed from whales can be identified to a fishery (Johnson et al., 2005). From 2010-2017, gear was retrieved from only 4% of confirmed entanglement cases (Knowlton et al., 2019). Likely ~2% of all confirmed entanglement cases therefore have identified gear. This is far too small of a proportion of the whole on which to attribute entanglement origin.
- Bias in gear identification: Again, due to the limited gear-marking scheme, many lines removed from entangled whales are nondescript portions of rope. The characteristics of these ropes vary (van der Hoop et al., 2016) but many cannot be traced back to their geographic or industrial origin. Conversely, small pieces of evidence can lead to the immediate identification of source for other fisheries. For example, a towed lobster trap (e.g., van der Hoop et al. 2016), a gillnet, and the unique distinguishing marks of Canadian snow crab gear allow for the rapid identification of these gear sets relative to unmarked line. This bias in the probability of identification given different fisheries will drive the proportion of cases where gear is identified to source.

NOAA has therefore proposed a 50/50 apportionment approach for unassigned serious injuries and mortalities between the US and Canada. This 50/50 apportionment results in similar numbers as the previous approach of assigning based on the location first-seen entangled.

There is concern that the 50/50 apportionment does not reflect the fraction of time the species spends in US waters, given the recent distribution shift. Prior to 2010, right whales likely spent >50% of time in US waters, but it is uncertain if the species is now spending less time in the US and an increasing amount of time in Canada, summer through fall. At least a third of the species' population are resident in Canada from at least June through October, with the Gulf of St Lawrence supporting ~130 unique whales per year (Crowe et al., 2019). The whereabouts of the majority of the species during these months is unknown, though the majority of the population is seen in US waters in spring (Ganley et al., 2019), and recent surveys have found right whales persisting in US waters through the summer and fall. Acoustic data suggest a year-round presence of right whales in US waters (Davis et al., 2017). A 50/50 split is likely conservative for the time spent in US waters, and the data on which to base a different level of apportionment do not exist.

Recommendations:

Given the known under-detection of entanglement events; the known delay in detecting such events; the ability for whales to move considerable distances in short times, even when entangled; and the small proportion of total entanglements that are observed, with gear that is obtained, and subsequently identified, there are too many factors that can be identified in the interest of any party involved in apportioning mortality (i.e., political, industry sector, etc.). Action, however, is required.

Location of first-seen entanglement is not a representative indicator of the origin of the entanglement, and gear identification cannot be relied upon. NOAA's proposed 50/50 apportionment is supported by regulation and precedent for PBR take calculations (Barlow et al., 1995), and still likely underestimates the proportion of time the species spends in US waters.

Conclusions

By definition, decision-support tools have six characteristics (Geoffrion, 1983):

i. explicit design to solve ill-structured problems
ii. easy-to-use and powerful user interface
iii. ability to combine analytical models with data
iv. ability to explore the solution space by building alternatives
v. capability of supporting a variety of decision-making styles
vi. allowing interactive and recursive problem-solving.

The DST evaluated here meets these six characteristics. It is explicitly designed to test combinations of multiple mitigation strategies over small or large scale across a large region, and the resulting effects on gear density, gear characteristics, threat to whales, and resulting risk given whale density in time and space. Its interface is easy to use for those comfortable executing R, though as a Reviewer I understand that a shiny-app is in the works for the future. I agree that this is not the top priority at the moment.

The DST combines diverse data sets and various model types to produce estimates used in its computation. It provides the ability to explore the solution space and allows end users to build their own mitigation strategies, supporting a variety of decision-making styles, cultures, expertise, knowledge, and experience. It allows for recursive problem-solving in allowing for iterations on design on relatively short timeframes.

In this way, the DST meets the requirements of providing a framework to identify realistic management choices, and allow for ownership and agency over proposed, evaluated, potential, and implemented management strategies. The transparency of the framework could be improved by providing clear documentation for all parameters, assumptions, and data used to create the tool itself. Also, the sensitivity of the DST results to these parameters and assumptions needs to be tested.

Decision support tools should allow agencies and stakeholders to weigh the costs of collecting and/or analyzing new data against the value of additional data in arriving at a more robust decision (Geoffrion 1983). I believe this a major opportunity that the DST and its component models has not taken advantage of. It is key to acknowledge the limitations and vulnerabilities of our methods, and to identify areas where data are needed and *exactly how* and *what types* of data would improve the quality of the models (and *by what percent*). Additionally, this can identify where our methods have developed conclusions we are confident in and where no additional data are needed to support any type of decision.

Decision support tools (in general) are designed to allow for uncertainty to be resolved through exploring the effects of alternative conceptual models and parameters choices on the decision, and serve to make the decision-making process transparent, documented, reproducible, robust and provide a coherent framework to explore the options available (Geoffrion 1983). This is a critical set of steps that have not yet been completed for the NARW entanglement DST. Results from a DST should be reproducible and distributed to a wide user base; however, the NARW entanglement DST has not been used to resolve uncertainties, and is not fully transparent or coherent in its framework and assumptions or those of its sub-components. The accuracy and sensitivity of the predictions are critical as they form the basis for environmental management and conservation (Li 2019).

Currently, we do not have the understanding of the consequence based on co-occurrence; to assign a risk reduction based on this transfer function is really not possible. The threat model component is not ready for prime time. However, it serves a true purpose in achieving the goal of engaging stakeholders to develop and test their own solutions for reducing lethal entanglement risk to whales.

This is not to say that the tool has to be improved prior to any decision-making. We already know that line in the water column poses risk, and that the numbers of lines and their soak times are considerably higher than needed. Also, a reduction in effort across the board may not reduce landings (Myers et al., 2007; Myers and Moore 2019).

A decision could be made and rules implemented while the next, coordinated, sensitivity tested iteration of the DST is completed. In three years, a decision could be informed by the full extent of the tool.

The unidentified serious injury and mortality 50/50 apportionment does not require any additional data or delay. This approach follows precedent, is defined by the law, and is diplomatic, allowing for immediate action.

References

Bardos, R., Mariotti, C., Marot, F. and Sullivan, T. (2001). Framework for decision support used in contaminated land management in Europe and North America. *NATO/CCMS Pilot Study*, 9.

Barlow, J., Brownell Jr, R. L., DeMaster, D. P., Forney, K. A., Lowry, M. S., Osmek, S., Ragen, T. J., Reeves, R. R. and Small, R. J. (1995). US Pacific marine mammal stock assessments.

Baumgartner, M. F. and Mate, B. R. (2003). Summertime foraging ecology of North Atlantic right whales. *Marine Ecology Progress Series* **264**, 123-135.

Baumgartner, M. F. and Mate, B. R. (2005). Summer and fall habitat of North Atlantic right whales (*Eubalaena glacialis*) inferred from satellite telemetry. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 527-543.

Baumgartner, M. F., Wenzel, F. W., Lysiak, N. S. J. and Patrician, M. R. (2017). North Atlantic right whale foraging ecology and its role in human-caused mortality. *Marine Ecology Progress Series* **581**, 165-181.

Carr, M. K., Vanderlaan, A. S. M., Davies, K. T. A. and Taggart, C. T. (2018). Relative vessel-strike risk to North Atlantic right whales in the Gulf of St. Lawrence. In *North Atlantic Right Whale Consortium*. New Bedford, MA, USA.

Cole, A. K. and Brillant, S. W. (2019). Decisions to implement spatio-temporal fisheries closures to reduce entanglement threats to whales must also consider how those closures can change the nature of a fishery. *NARW Consortium Meeting Abstract Booklet*.

Crowe, L. M., Brown, M., Corkeron, P., Duley, P., Hamilton, P., Ogilvie, A., Ratelle, S., Vanderlaan, A. S. M. and Cole, T. V. N. (2019). An update on the population structure, residency, and movements of North Atlantic right whales in the Gulf of St. Lawrence. *NARW Consortium Meeting Abstract Booklet*.

Davies, K. T. A. and Brillant, S. W. (2019). Mass human-caused mortality spurs federal action to protect endangered North Atlantic right whales in Canada. *Marine Policy* **104**, 157-162.

Davies, K. T. A., Taggart, C. T. and Smedbol, R. K. (2014). Water mass structure defines the diapausing copepod distribution in a right whale habitat on the Scotian Shelf. *Marine Ecology Progress Series* **497**, 69-85.

Davies, K. T. A., Vanderlaan, A. S. M., Smedbol, R. K. and Taggart, C. T. (2015). Oceanographic connectivity between right whale critical habitats in Canada and its influence on whale abundance indices during 1987–2009. *Journal of Marine Systems* **150**, 80-90.

Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C., Bort Thornton, J., Brault, S., Buchanan, G., Charif, R. A., Cholewiak, D. et al. (2017). Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. *Sci Rep* **7**, 13460.

Frasier, T. R., Hamilton, P. K., Brown, M. W., Conger, L. A., Knowlton, A. R., Marx, M. K., Slay, C. K., Kraus, S. D. and White, B. N. (2007). Patterns of male reproductive success in a highly promiscuous whale species: the endangered North Atlantic right whale. *Molecular ecology* **16**, 5277-93.

Ganley, L. C., Brault, S. and Mayo, C. A. (2019). What we see is not what there is: estimating North Atlantic right whale Eubalaena glacialis local abundance. *Endangered Species Research* **38**, 101-113.

GAO. (2009). National Marine Fisheries Service: improvements are needed in the federal process used to protect marine mammals from commercial fishing, vol. GAO-09-78 (ed. G. A. Office). Washington, DC.

Geoffrion, A. M. (1983). Can MS/OR evolve fast enough? *Interfaces* **13**, 10-25.

Johnson, A., Salvador, G., Kenney, J., Robbins, J., Kraus, S., Landry, S. and Clapham, P. (2005). Fishing gear involved in entanglements of right and humpback whales. *Marine Mammal Science* **21**, 635-645.
Knowlton, A. R., Marx, M. K. and Pettis, H. M. (2019). An update on right whale entanglement interactions: 2010-2019. *NARW Consortium Meeting Abstract Booklet*.
Knowlton, A. R., Robbins, J., Landry, S., McKenna, H. A., Kraus, S. D. and Werner, T. (2016). Implications of fishing rope strength on the severity of large whale entanglements. *Conservation Biology* **30**, 318-328.
Li, J. (2019). A Critical Review of Spatial Predictive Modeling Process in Environmental Sciences with Reproducible Examples in R. *Applied Sciences* **9**, 2048.
Malik, S., Brown, M. W., Kraus, S. D., Knowlton, A. R., Hamilton, P. K. and White, B. N. (1999). Assessment of mitochondrial DNA structuring and nursery use in the North Atlantic right whale (*Eubalaena glacialis*). *Canadian Journal of Zoology* **77**, 1217-1222.
Mayo, C. and Goldman, L. (1992). Right whale foraging and the plankton resources in Cape Cod and Massachusetts Bays. In *The right whale in the western North Atlantic: a science and management workshop (ed. J. Hain)*, pp. 43-44.
Mayo, C. A., Letcher, B. H. and Scott, S. (2001). Zooplankton filtering efficiency of the baleen of a North Atlantic right whale, Eubalaena glacialis. *Journal of Cetacean Research and Management* **(Special Issue) 2**, 225-229.
Mayo, C. A. and Marx, M. K. (1990). Surface foraging behaviour of the North Atlantic right whale, *Eubalaena glacialis*, and associated zooplankton characteristics. *Canadian Journal of Zoology* **68**, 2214-2220.
Moore, M. J., van der Hoop, J. M., Barco, S. G., Costidis, A. M., Gulland, F. M., Jepson, P. D., Moore, K. T., Raverty, S. and McLellan, W. A. (2013). Criteria and case definitions for serious injury and death of pinnipeds and cetaceans caused by anthropogenic trauma. *Diseases of Aquatic Organisms* **103**, 229-264.
Murison, L. and Gaskin, D. E. (1989). The distribution of right whales and zooplankton in the Bay of Fundy, Canada. *Canadian Journal of Zoology* **67**, 1411-1420.
Myers, R. A., Boudreau, S. A., Kenney, R. D., Moore, M. J., Rosenberg, A. A., Sherrill-Mix, S. A. and Worm, B. (2007). Saving endangered whales at no cost. *Current Biology* **17**, R10-R11.
NMFS. (2012). Process for distinguishing serious from non-serious injury of marine mammals. *National Marine Fisheries Service Policy Directive PD 02-038. Department of Commerce, National Oceanic and Atmospheric Adminstration, National Marine Fisheries Service.*, 4 pp. Available at: http://www.nmfs.noaa.gov/directives/.
Nowacek, D. P., Johnson, M. P. and Tyack, P. L. (2004). North Atlantic right whales (*Eubalaena glacialis*) ignore ships but respond to alerting stimuli. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **271**, 227-31.
Pace, R. M., 3rd, Corkeron, P. J. and Kraus, S. D. (2017). State-space mark-recapture estimates reveal a recent decline in abundance of North Atlantic right whales. *Ecol Evol* **7**, 8730-8741.
Parks, S. E., Searby, A., Célérier, A., Johnson, M. P., Nowacek, D. P. and Tyack, P. L. (2011). Sound production behavior of individual North Atlantic right whales: implications for passive acoustic monitoring. *Endangered Species Research* **15**, 63-76.
Pettis, H. M., Rolland, R. M., Hamilton, P. K., Knowlton, A. R., Burgess, E. A. and Kraus, S. D. (2017). Body condition changes arising from natural factors and fishing gear entanglements in North Atlantic right whales *Eubalaena glacialis*. *Endangered Species Research* **32**, 237-249.
Roberts, J. J., Best, B. D., Mannocci, L., Fujioka, E., Halpin, P. N., Palka, D. L., Garrison, L. P., Mullin, K. D., Cole, T. V. N., Khan, C. B. et al. (2016). Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Sci Rep* **6**, 22615.

Rolland, R. M., Schick, R. S., Pettis, H. M., Knowlton, A. R., Hamilton, P. K., Clark, J. S. and Kraus, S. D. (2016). Health of North Atlantic right whales *Eubalaena glacialis* over three decades: from individual health to demographic and population health trends. *Marine Ecology Progress Series* **542**, 265-282.
Schick, R. S., Kraus, S. D., Rolland, R. M., Knowlton, A. R., Hamilton, P. K., Pettis, H. M., Kenney, R. D. and Clark, J. S. (2013). Using hierarchical Bayes to understand movement, health, and survival in the endangered North Atlantic right whale. *PloS one* **8**, e64166.
Sharp, S. M., McLellan, W. A., Rotstein, D. S., Costidis, A. M., Barco, S. G., Durham, K., Pitchford, T. D., Jackson, K. A., Daoust, P.-Y. and Wimmer, T. (2019). Gross and histopathologic diagnoses from North Atlantic right whale Eubalaena glacialis mortalities between 2003 and 2018. *Diseases of Aquatic Organisms* **135**, 1-31.
van der Bles, A. M., van der Linden, S., Freeman, A. L. J., Mitchell, J., Galvao, A. B., Zaval, L. and Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science* **6**, 181870.
van der Hoop, J. M., Corkeron, P., Kenney, J., Landry, S., Morin, D., Smith, J. and Moore, M. J. (2016). Drag from fishing gear entangling North Atlantic right whales. *Marine Mammal Science* **32**, 619-642.
van der Hoop, J. M., McGregor, A. E. N., Nowacek, D. P., Parks, S. E., Tyack, P. L. and Madsen, P. T. (2019). Foraging rates of ram-filtering North Atlantic right whales measured with onboard, multi-sensor DTAGs. *Functional Ecology* **33**, 1290-1306.
Vanderlaan, A. S. M., Smedbol, R. K. and Taggart, C. T. (2011). Fishing-gear threat to right whales (*Eubalaena glacialis*) in Canadian waters and the risk of lethal entanglement. *Canadian Journal of Fisheries and Aquatic Sciences* **68**, 2174-2193.
Vanderlaan, A. S. M., Taggart, C. T., Serdynska, A. R., Kenney, R. D. and Brown, M. W. (2008). Reducing the risk of lethal encounters: vessels and right whales in the Bay of Fundy and on the Scotian Shelf. *Endangered Species Research* **4**, 283-297.
Wishner, K. F., Schoenherr, J. R., Beardsley, R. and Chen, C. (1995). Abundance, distribution and population structure of the copepod Calanus finmarchicus in a springtime right whale feeding area in the southwestern Gulf of Maine. *Continental Shelf Research* **15**, 475-507.

Appendix 1: Bibliography of materials provided for review

NMFS, 2019. *CIE Review Background: Take Reduction Target.* Greater Atlantic Regional Fisheries Office, Protected Resources Division briefing document.

NMFS, 2019. *CIE Review Background: Decision Support Tool Model Documentation.* NOAA Fisheries Northeast Fisheries Science Center, Protected Resources Division briefing document.

Appendix 2:

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

North Atlantic Right Whale Decision Support Tool

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

NMFS is required to use the best available scientific and commercial data in making determinations and decisions under the Endangered Species Act (ESA) and Marine Mammal Protection Act (MMPA). Right whales, humpback whales, and fin whales are listed as endangered species under the ESA. Pursuant to the ESA and the MMPA, the National Marine Fisheries Service (NMFS) – with guidance from the Atlantic Large Whale Take Reduction Team (ALWTRT) – is responsible for the development and implementation of measures to reduce the risks of entanglement. These measures are embodied in the Atlantic Large Whale Take Reduction Plan (ALWTRP). The plan seeks to reduce the risks of entanglement through a set of gear modifications and other requirements that affect commercial fishing operations in Atlantic waters.

A continuing concern in the evolution of the ALWTRP is the risk of entanglement in vertical line; i.e., buoy lines associated with lobster trap/pot gear, or other fixed gear. To better understand this risk and, particularly, the potential impact of management measures designed to address it, NMFS requires information on the risks of entanglement and injury associated with vertical line used by various fisheries amount of vertical line used by various fisheries, especially the extent to which that line is fished in areas and during seasons in which whales are likely to be present. An absolute measure of entanglement risk is not feasible, but measures of relative risk are possible. At the most recent ALWTRT meeting in April 2019, NMFS introduced a North Atlantic Right Whale Decision Support Tool (DST) to help understand relative risk of entanglement in different geographic locations, and, most importantly, the reduction in relative risk based on different proposed mitigation scenarios.

The information and analysis contained in the report to be presented will include essential factual elements upon which the agency may base its rule-making determination. Accordingly, it is critical that the reports contain the best available information on the relative risk and reduction in relative risk based on mitigation scenarios, and that all scientific findings be both reasonable and supported by valid information contained in the documents. Therefore, the CIE reviewers will conduct a peer review of the scientific information and mathematical approach in the DST based on the Terms of Reference (ToRs). The CIE reviewers will ensure an independent, scientific review of information for a management process that is likely to be controversial.

The specified format and contents of the individual peer review reports are found in **Annex 1**. The specified format and contents of the summary report are found in **Annex 2**. The Terms of Reference (ToRs) for the review of the North Atlantic Right Whale DST are listed in **Annex 3**. Lastly, the tentative agenda of the panel review meeting is attached in **Annex 4**.

Requirements
NMFS requires three reviewers to conduct an impartial and independent peer review in accordance with the PWS, OMB guidelines, and the TORs below. The reviewers shall have a working knowledge and recent experience in the application of one or more of the following: 1) Atlantic large whales and entanglement; 2) Co-occurrence risk modeling; 3) Fixed gear/fishing rope strength and the severity of whale entanglements; 4) Lethal and sublethal impacts of interactions with fishing gear on protected species.

Tasks for Panel Reviewers

1) Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewer the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to

the reviewer in accordance to the PWS scheduled deadlines specified herein. Each CIE reviewer shall read all documents in preparation for the peer review.

Background documents will be provided by NMFS prior to the CIE review.

2) Panel Review Meeting: The CIE reviewers shall conduct the independent peer review in accordance with the PWS and ToRs, and shall not serve in any other role unless specified herein. Modifications to the PWS and ToRs cannot be made during the peer review. The CIE reviewers shall actively participate in a professional and respectful manner as members of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

3) Contract Deliverables - Independent CIE Peer Review Report: The CIE reviewers shall complete an independent peer review report in accordance with the PWS. The CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. The CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

4) Other Tasks – Contribution to Summary Report: The CIE reviewers will assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of their views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

5) Deliver their reports to the Government according to the specified milestones dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration- system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The place of performance shall be at the contractor's facilities, and at the Northeast Fisheries Science Center in Woods Hole, MA.

Period of Performance
The period of performance shall be from the time of award through January 2020. The CIE reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers' participation
At least two weeks prior to the panel review meeting	Contractor provides the pre-review documents to the reviewers
November 19-21, 2019	Each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks after review	Contractor receives draft reports and summary report
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards:
(1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each ToR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $10,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:

Tara Trinko Lake
NMFS/Northeast Fisheries Science Center
166 Water St.
Woods Hole, MA 02540
508-495-2395
tara.trinko@noaa.gov

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Performance Work Statement
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the chair that will include the background and a review of activities and comments on the appropriateness of the process in reaching the goals of the review. Following the introduction, the report should address whether or not each Term of Reference of the Right Whale Decision Support Tool review was completed successfully. For each Term of Reference, the Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing management advice. If the reviewers and chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

   The report may include recommendations on how to improve future use of the Right Whale Decision Support Tool.

2. The report shall also include the bibliography of all materials provided during the review, and relevant papers cited in the Summary Report, along with a copy of the CIE Statement of Work.

Annex 3: Terms of Reference
For the North Atlantic Right Whale Decision Support Tool

1. Evaluate the data inputs (e.g., spatial and seasonal gear configuration, spatial and seasonal right whale distribution, etc.) used in the Decision Support Tool.
2. Evaluate the data outputs (e.g., vertical line estimates, relative risk to right whales, etc.) produced by the Decision Support Tool.
3. Comment on the appropriateness of using the Decision Support Tool as an approach to evaluate relative entanglement risk to right whales and advise on the strengths and weaknesses of using the DST to compare management measures. The goal is to understand the relative risk of entanglement in different geographic locations and the reduction in relative risk based on different proposed mitigation scenarios.
4. Provide research recommendations for further improvement of the Decision Support Tool.
5. Evaluate whether the methods represent the best available scientific approach for apportioning anthropogenic mortality by country.

Annex 4: Tentative Agenda – Panel Review

North Atlantic Right Whale Decision Support Tool

Woods Hole, MA
November 19-21, 2019

North Atlantic Right Whale Decision Support Tool

Woods Hole, MA
November 19-21, 2019

Tuesday, November 19, 2019

Time	Activity	Lead
10:00 am	Welcome and Introductions	Sean Hayes/Tara Trinko Lake
10:10 am	Overview and Process	Sean Hayes/Tara Trinko Lake
10:30 am	TRT Background **[Coogan PPT 1]**	Mike Asaro/Colleen Coogan
11:00 am	Co-Occurrence Model- **[Etre PPT 1]**	IEC Neil Etre
11:30 am	Decision Support Tool Purpose and Scope **[Hayes PPT 1]**	Sean Hayes
11:45 am	Model Overview and Fishery Inputs **[Shank PPT 1]**	Burton Shank / IEC
12:15 pm	**Lunch**	
1:15 pm	Fishery Inputs Continued	Burton Shank
2:00 pm	Discussion/ Review of Fishery Inputs	Review Panel
2:30 pm	Model Inputs: Gear Threat **[Shank PPT 2]**	Burton Shank / PSB Staff
3:15 pm	**Break**	
3:30 pm	Model Inputs: Gear Threat Continued	Burton Shank / PSB Staff
4:15 pm	Discussion / Review of Gear Threat Model	Review Panel
4:45 pm	Public Comment	Public
5:00 pm	General Discussion / Day1 Wrap-up	Review Panel / Presenters
5:30 pm	Adjourn	

Wednesday, November 20, 2019

Time	Activity	Lead
9:00 am	Brief Overview and Logistics	Sean Hayes/ Tara Trinko Lake
9:10 am	Model Inputs - Whale Habitat Modeling **[Roberts PPT 1]**	Jason Roberts
10:30am	Discussion / Review of Whale Habitat Modeling	Review Panel
11:00pm	Public Comment	Public
11:15 am	**Break**	
11:30am	Model Inputs- User Configurations	Burton Shank

12:30 pm	**Lunch**	
1:30 pm	Discussion / Review of User Inputs	Review Panel
1:45 pm	Model outputs- Risk to Right Whales	Burton Shank
2:45 pm	**Break**	
3:00 pm	Model Outputs- Risk to Right Whales	Review Panel
Discussion/Review/Summary		
4:15 pm	Public Comment Public	
4:30 pm	General Discussion/Day 2 Wrap-Up	Review Panel/
Presenters		
Key Topics		
5:00 pm	Adjourn	
Thursday, November 20, 2019		
9:00 am	Brief Overview and Logistics	Sean
Hayes/Tara Trinko Lake		
9:10 am	Right Whale Mortality Apportionment **[Coogan PPT 2]**	Colleen
Coogan		
10:10 am	Discussion/Review of Mortality Apportionment	Review Panel
10:40 am	Public Comment	Public
10:55 am	Break	
11:10 am	Meeting Wrap-Up and Discussion of Key Topics	Review Panel
12:00 pm	Lunch	
1:00 pm	Report Writing	Review Panel
5:00 pm	Adjourn	

*All times are approximate, and may be changed at the discretion of the chair. The meeting is open to the public; however, during the Report Writing sessions we ask that the public refrain from engaging in discussion with the reviewers.

Appendix 3: Panel membership

The review panel consisted of Dr. Julie van der Hoop (Independent), Dr. Jason How (Department of Primary Industries and Regional Development, Western Australia) and Dr. Don Bowen (Dalhousie University).

Appendix 4: Outputs from the Decision Support Tool

Model Outputs

1. Low-Resolution Monthly Maps of the following Default conditions:

1.1. Trap density
1.2. Mean trawl length
1.3. Vertical line density
1.4. Mean vertical line strength
1.5. Mean gear threat score
1.6. Total threat score (gear threat * line density)
1.7. Whale density
1.8. Total risk (total threat * whale density).

An .Rdata file with the individual data objects used for creating these maps is also saved to custom maps can be created after the model run.

2. Low-Resolution Monthly Maps of the following Scenario conditions:

2.1. Trap density before scenario effects on traps
2.2. Trap density after trap reduction
2.3. Trap density after implementation of trap caps
2.4. Trap density after implementation of closures
2.5. Map of traps relocated as a result of closures
2.6. Trawl lengths after scenario effects
2.7. Line densities after scenario effects
2.8. Mean line strength after scenario effects
2.9. Mean gear threat after scenario effects
2.10. Total gear threat after scenario effects
2.11. Whale densities
2.12. Total risk scores.

An .Rdata file with the individual data objects used for creating these maps is also saved to custom maps can be created after the model run.

3. Output tables

3.1. Model documentation
3.1.1. Model configuration settings
3.1.2. Contents of the input spreadsheet

These two outputs allow users to fully understand the settings of a model run as well as recreate the model run a later time.

3.2. Tables with monthly values for default and scenario conditions
3.3. Initial and final trap numbers
3.4. Total number of trawls

3.5. Mean trawl length
3.6. Total vertical lines
3.7. Mean vertical line strength
3.8. Mean threat score per vertical line
3.9. Total gear threat
3.10. Seasonal whale density
3.11. Total risk scores

All summary statistics written to the tables are also written to a comma-separated text file for further access.